

Yes, They Can All Succeed:

Implementation of Standards-Based Assessments in the United States for Primary and Secondary School Students*

SCOTT LIU[†]

University of Illinois at Urbana-Champaign
sliu125@illinois.edu

April 16, 2016

Abstract

Although grading is generally acknowledged to be a universal technique for student evaluation, the various methods of grading itself are far less known amongst students and teachers alike. This paper illustrates the common misconceptions concerning Standards-Based Assessments and also provides a comparison between Standards-Based Assessments and other student grading methods. Researchers have repeatedly pointed out the inconsistent use of terminology regarding assessment methods. Through the analysis of many journals in this field of research, it is evident that concepts such as “analytic scoring” and “holistic scoring” should be rigorously defined. As a result, I have developed clear definitions for such concepts for reference of future academics. On an international scale, the United States supported a standards-based education reform after the Elementary and Secondary Education Act [No Child Left Behind] in 2001. This paper analyzes statistics from the U.S. Department of Education and compares cases of successful applications of Standards-Based Assessments worldwide. Finally, I conclude that statewide assessments in the U.S. should be based on the Standards-Based Assessments alone for elementary and secondary school students to better prepare them for the post-secondary environment.

I. KEYWORDS

Standards-Based Assessment; Evaluation Methods; Academic Standards; Grading; Secondary Education

II. INTRODUCTION

There is no doubt that education will forever remain core in the human civilization. As knowledge is acquired, our understanding in that specific domain becomes broadened. While we may not all be epistemologists ourselves, many of us are still involved in the field of education one way or another. It is therefore in our best interest to develop a practical method that incorporates both learning objectives and learning outcomes

while connecting the two in an efficient, cause-effect relationship. In the general student assessment context, many solutions have been proposed throughout the years. These proposals, also known as grading methods, include but are not limited to: Standards-Based Assessments (SBA), Criterion-Referenced Tests (CRT), Norm-Referenced Tests (NRT), holistic scoring, and analytic scoring. On one hand, Standards-Based Assessments can be considered as an extension from Criterion-Referenced Tests with a surprisingly large number of similarities; but on the other hand, they have subtle yet significant differences which will be covered and elaborated on in a later section. Despite the general consensus agreeing that the education reform in the U.S. – using a multitude of grading practices – is already successful enough as it currently stands, this paper is purposed to argue that statewide assessments in the U.S. should be based on the Standards-Based As-

*Written as a semester-long research assignment for RHET 105

[†]Also known as Sicong Liu

assessments alone for elementary and secondary school students because the key improvements, namely more accessible documentation of standards and more adequate feedback to students and their guardians, are more suited to prepare them for the post-secondary environment.

The curious reader may ask why the United States was chosen to be the context upon which this article's argument was based. The answer lies not only within the personal experience of its author, but also in the well documented education reform in America [No Child Left Behind]. As an international undergraduate studying at the University of Illinois at Urbana-Champaign with high school qualifications from New Zealand, I bear a unique position that allows me to comment on this new educational system here based on my prior experience with a different system in which Standard-Based Assessments take place [New Zealand Qualification Authority #3]. Additionally, it has come to my attention that grades are occasionally curved depending on how well the entire class has done throughout a semester; while some courses include free response as an assessment method and grade this based on a rubric, some simply only offer multiple choice questions in the exam and grade everything using electronic devices. This partially reflects on the current chaotic alignments followed by states in the U.S., with almost a different assessment type for every state [U.S. NCES], each with small variations, but mostly still based on the Criterion-Referenced Test. Furthermore, an educational reform brings change to the system itself. This gives me the opportunity to analyze data before and after to produce meaningful results regarding how effective this reform has been.

In the discussion regarding holistic scoring and analytic scoring, while some academics have only hinted about errors in certain existing definitions [Hamp-Lyons], others have been angered by an inconsistent use of terminology in grading related articles [Harsch and Martin]. Although a slight deviation from an exact concept may seem trivial,

it is in fact detrimental when it comes to making academic arguments because it defeats the purpose of being scholarly when two parties are disputing about entirely different concepts. Conversely, the use of consistent terminology would be extremely beneficial and would lead to constructive reasoning and hence further the cause. In the next section, I address this issue by first analyzing how each student evaluation method works and then making appropriate comparisons to see the various similarities and differences between them. This process is important in order to fully comprehend the characteristics of each technique. To truly resolve this issue, a new set of definitions must be given to act as a guideline for future scholars in this field. Given that there is such a need for definitions, this necessary piece of information shall follow the comparison. The bulk of the paper is focused on the analysis of Standards-Based Assessment itself and data from United States Department of Education, followed and ended by a brief conclusion to summarize my findings and reinforce my main points.

III. CATEGORIZATION

Standard-Based Assessments, Norm-Referenced Tests, holistic scoring, and analytic scoring are contrasted to show the fundamental underlying connection between these four concepts¹. These links have been overlooked in the current ongoing debate. In fact, many disputes such as ones caused by the misunderstanding of these concepts themselves can be resolved by closely analyzing these relationships.

While Standards-Based Assessments references students' achievement against a set of predetermined standards [Sharma], Norm-Referenced Tests assess students by, as the name would imply, comparing them to the norm; this is also commonly called grading on the curve [Aviles 604]. As an analogy, consider the score from a Standard-Based Assessment to reflect the percentage an individual

¹NRT is not categorized here as it is a more general form of SBA.

answered correctly out of the total; the score from a Norm-Referenced test would reflect how many more (or less) questions an individual answered correctly relative to their peers. Additionally, it is important to note that in a Standards-Based environment, students should be well aware of these standards and know what is expected from them [Sharma]. This is the most evident property between the two groups, and it will be evaluated extensively because it gives rise to more significant consequences than other factors. Even though the academic conversation implies that holistic scoring fits in the side where the results are graded based on personal standards only [Schullery], I must still contend that the opposite could be true nonetheless, as the grades from these types of grading could just as easily be subject to peer comparison. Therefore what the results are actually compared to really just vary from situation to situation.

Another important aspect we can categorize by is determining whether the methods relies on rubrics or not. It is generally agreed that holistic scoring does not use a rubric and analytic scoring depends heavily on a rubric [Schullery]. However, this gets more complicated when we take Standards-Based Assessments and Norm-Referenced Tests into account. In the case for Standards-Based Assessments, Sharma summarizes that assessment tasks are scored on the basis of rubrics derived from Standards themselves (12). Although I agree with her wholeheartedly, I would also like to add that these rubrics are not to be followed blindly; instead, the grader should use it as a general guideline and apply professional judgment when necessary much like that in holistic scoring, even though holistic scoring doesn't use rubrics at all. Given that Norm-Referenced Tests are only characterized by a normalized final result, whether it uses rubrics or not is completely determined by the type of the test itself. This leads to a classification of assessment methods by test types too.

As a rule of thumb, Standard-Based Assessments generally use free response due to the nature of the results being graded against a

standard. A major discrepancy occurs when we arrive at the cases of holistic and analytic scoring. Hamp-Lyons rules out that analytic scoring should be (and only be) used for multiple choice types [Hamp-Lyons]. Although it is tempting to immediately agree with her claim, as her simple resolution seems to be effective, it does seem unjustified without any direct evidence, considering that some people would classify mixed-response type also as analytic. After being puzzled for a brief moment, we quickly realize that these concepts themselves are loosely defined and this is what's causing the confusion. This accentuates the importance of giving clear-cut definitions about holistic and analytic scoring in the next section. For now, we will just stick with the holistic method using free response only and the analytic method using multiple choice only with nothing in between.

Feedback is one of the trademark characteristics of Standard-Based Assessments [Sharma]. In all articles regarding this topic, feedback is regarded as a highly important feature that sets Standard-Based Assessments apart from the others. Along the same lines, feedback is also usually given during holistic scoring. This is especially clear in Schullery's case, where she explicitly states how she gives positive reinforcement to the class (89). Nevertheless, she does also emphasize about giving no personalized feedback due to her belief that she shouldn't make anyone feel special (89). Although I interpret her intent as being benevolent, I must question her assumption that personalized feedback could only be given to a single student. Instead, an alternative approach could be adopted where personalized feedback was given to all students. In both Norm-Referenced Tests and analytic scoring however, there has been no evidence to show that either of these two methods provide any feedback to the student at all. Assuming that feedback provides a positive effective to students (the details will be analyzed later on), it necessarily follows that if we include feedback into the aforementioned two methods, they can be considered a more effective evaluation

Table 1: Comparison between assessment types

	Rubrics	Test type	Feedback
SBA	Standard	FR	Important
NRT	Depends	Many	Not given
Holistic	None	FR	Important
Analytic	Heavily	MC	Not given

method.

To summarize the previous comparison succinctly, table 1 below provides a brief overview on what has been covered already.

IV. DEFINITIONS

- **Standard-Based Assessment² (SBA):** A method of student evaluation that measures students' proficiency level based on preset, well defined standards. Students should have ready access to sources that inform about what is expected from them (the standards). The result from the evaluation process should be clearly communicated to the student, their guardians, and other teachers.
- **Criterion-Referenced Test (CRT):** A method of student evaluation that measures students' proficiency level by referencing to preset, well defined standards.
- **Norm-Referenced Test (NRT):** A type of student evaluation that yields an estimate of their relative position with respect to some population. This result is usually reported as a percentage or percentile ranking.
- **Holistic scoring:** Any grading method that only involves free response questions. It is commonly graded based on professional judgement, and feedback is usually given.
- **Analytic scoring:** Any grading method that only involves multiple choice questions. It is graded with strict reference to a rubric, and usually no feedback is given.

²The SBA can be considered as a special case of the CRT. All references of CRT in this paper will include SBA as a special case by default unless otherwise stated.

V. DISCUSSION

The effectiveness of certain grading methods have been proposed to be somewhat controversial. Some would even go as far as to argue if we should grade students at all [C. Edwards and L. Edwards]. A study in 2011 offers data regarding the extent to which teachers follow certain principles [Tierney et al.]. The results clearly suggested the principle of Criterion-Referenced Testing was not at all controversial, with more than 93 percent participants agreeing. It is an indication that we still hold firm beliefs on the effectiveness of Criterion-Referenced Tests in general. Knowing that Standards-Based Assessments is a stricter manifestation of the Criterion-Referenced, the fact that Standards-Based Assessments would be a highly regarded effective evaluation method is indirectly demonstrated.

Having just argued about the potential in Standard-Based Assessments, let us focus on some aspects that are unique to this method itself. Extensive feedback is a main feature in Standards-Based Assessments. Sharma claims that the feedback provided to students by their teachers is crucial, enabling them to advance their learning [Sharma]. I would like to build upon her argument by also stressing the importance of feedback provided to their guardians. Moreover, I believe that it is best to keep the guardians informed of their child's academic progress, whether they are ahead of schedule or falling behind. This gives the guardian more knowledge and freedom to make strategic decisions suitable for their child. Although many teachers do communicate grades to parents, I found that teachers were considerably less likely to explain the grades in detail to parents compared to students by analyzing a table in a research journal [Tierney et al.]. The effectiveness of positive feedback has also been investigated by Hattie and Timperley, who in their meta-analysis reported on not only the type of feedback we mentioned in Standards-Based Assessments, but also other types of feedback with a large number of parameters. They concluded that

when paired with effective instruction in the classroom, feedback is very effective in enhancing learning [Hattie and Timperley]. In sum, then, the act of providing adequate feedback in Standards-Based Assessment makes it a more effective evaluation method than the Criterion-Referenced method in this particular perspective.

Another feature of Standards-Based Assessment is how it values professional judgement from the grader as a backup method when ambiguity exists between the student's answer and the standard [Tierney et al.]. It is often assumed that grading Standards-Based Assessments requires less personal experience compared to grading a Norm-Referenced Test. This assumption cannot be justified, as understanding the standards correctly already require a substantial amount of teaching experience. Tierney observes that less experienced teachers focused on being fair than actually grading strictly based on the standard. While I would argue that being fair is also a form of professional judgement, he proposes one promising possibility to address this issue – by focusing more on providing the graders a better understanding of what the standards actually meant. Let me clarify that he is not expressing that the standards themselves are unclear, but simply asserting that less experienced teachers could still benefit from a better understanding.

So what do people think about the Norm-Referenced Test, which grades students on a curve? Aviles, a social work educator, would certainly take a strong position. In fact, he is ready to challenge his colleagues to abandon the practice of Norm-Referenced grading (608). The author attacks Norm-Referenced grading by arguing that it can mask teaching deficiencies and makes it difficult to measure any improvement in teaching skills (606). He also give readers a lively example of a student scoring well in a test but still receiving an unexpected grade. In the end, however, he does admit that Criterion-Referenced grading may indeed be less competitive in comparison and ultimately suggests that there are situations where Norm-Referenced measurements prevail and also sit-

uations where Criterion-Referenced measurements make better sense (607). Similarly, even though Sharma does comment on the advantages of Norm-Referencing, reminding us that there is simplicity when one interprets the scores; however, she also calls it out on the same idea, insisting that these simple scores fail to provide any meaningful information about what the student knows (11). As an extension from what Sharma claimed, we can interpret another group of scholars' conclusions to be an emphasis on what Standards-Based Assessment is capable of – it provides better feedback for the students and prepares them for post-secondary education [Shippy et al. 16].

In order to tie my hypothesis and findings with real world implications in the United States, it is necessary to have a good understanding about any major changes the contemporary education department has undergone. As a result, it is not hard to find the standards-based education reform after the Elementary and Secondary Education Act [No Child Left Behind] in 2001 being the most significant of all. The motivation of this education reform was based on the concern over the achievement of U.S. students on a global scale [U.S. NCES]. Among other categories, one major focus for the education reform was updating “standards, assessments, and accountability” (2). We can analyze how effective this process has been by researching relevant academic data before and after this reform. Analysis of a grading system table found in an old article published in 1994 showed that 37.2% of schools used a Criterion-Referenced system, and that 11.2% used a Normative-Referenced system [Polloway et al.], though it remains unclear whether these results were simply generalized across all subject or from one particular subject. On the other hand, data from the U.S. Department of Education in 2006 show that out of the 51 states investigated, 48 of them have been using Criterion-Referenced tests for the subject areas “English and Language arts” and “Mathematics”. With the goal being the enforcement of better standards across all states in the U.S., given that the percentage of Criterion-

Referenced test has gone up by a large margin, the result presented is a good indication that standards-based education reform has been a success.

Despite the clear lack of data from the National Center of Educational Statistics regarding the change in number for implementations of Standards-Based Assessments and the corresponding results, a select few samples were still retrieved which partially reflected how Standards-Based Assessments suited the new state expectations. Alaska was the only state that truly implemented Standards-Based Assessment. It was used to replace the Benchmark exams for grades 3 through 9. In contrast, Delaware had exams consisting of a Criterion-Referenced component as well as a Norm-Referenced component. The two reports both included a standard error that measures how closely their assessment methods map to the state expectations. What's really interesting is that the two states both demonstrated similar low error values, which may be an indication that these assessment methods both map well to the state expectation. Now don't get me wrong here, I'm not at all discrediting Standards-Based Assessments by showing how it produces similar results to this hybrid testing method. Instead, I'm trying to demonstrate that even though Standard-Based Assessment does have strong competitors, it still performs at an exceedingly high level. Given that these reports are from different states entirely, even when the results are similar, we still cannot be too certain when determining the better candidate. Undocumented sources of error may include differences in state expectations, differences in how well these techniques were executed, or even differences in educational funding; all of which may have unexpected effects on the case.

Other countries have also attempted to implement Standards-Based Assessments after realizing the benefits this evaluation method could potentially bring to their educational system. First phasing in the National Certificate of Educational Achievement (NCEA) in 2002 [Swarbrick 4], New Zealand uses a standard-

based system in which assessment standards describe what students have to achieve to gain credits. Different to other grading systems, achievement standards in New Zealand allow students to gain one of four grades – Not Achieved, Achievement, Achievement with Merit, and Achievement with Excellence (NCEA). This shows that a standards-based system can still be very flexible in terms of changing the details in a justified manner, when applied to a specific environment. On the other hemisphere, Sharma has tried hard to also bring Standards-Based Assessments to the Oriental classrooms. She mainly argues that Standard-Based Assessments are worth implementing in India because of the various benefits it can provide to students and teachers (27). She draws some key issues to the table that need to be addressed before the implementation of Standards-Based Assessments (26), all of which should be thought through thoroughly when considering to initially implement Standards-Based Assessment. Given that the context in which I propose Standards-Based Assessment has already been through an education reform many years ago, these considerations should not need to be explicitly addressed. With Criterion-Referenced tests already being the norm for fulfilling state standards in the U.S., we may safely assume that local educators have already been proficient in the measurement of student abilities based on preset standards or guidelines. The transition from Criterion-Referenced to Standards-Based therefore only faces to challenges: making the documentation of standards readily available to all students; and providing adequate feedback to students and effective communication to their guardians.

VI. CONCLUSION

In the perspective of general scholars, the purpose of this article should be to inform. Through the categorization of four different but related concepts, we recapitulate that Standards-Based Assessments evaluate students' answers based on preset standards.

Norm-Referenced Tests provide students with an estimate of the relative level they are amongst their peers. Additionally, rubrics (in different forms) are generally used in all of these methods except for holistic scoring. Furthermore, analytic scoring is defined to only include multiple-choice questions; holistic scoring only includes free response. Feedback is most important in Standards-Based Assessment, while holistic scoring also commonly includes feedback, it may be less personalized. Finally, Norm-Referenced and analytic scoring seldom include any feedback at all.

In the perspective of an expert in the field of education, I dedicate my reasoning to earnestly persuade. Through my series of justification for the cause of Standards-Based Assessments, I aim to present solid evidence that the addition of documentation and feedback will be constructive in terms of re-introducing this framework upon which we evaluate students. By bringing my own experience alongside specific cases of Standard-Based Assessment implementation, new insight is brought to the field and our understanding is refreshed. In the U.S. context, given that Criterion-Referenced tests has already been common practice, I advocate for more Standard-Based Assessments to take place for elementary and secondary school students to better prepare them for the post-secondary environment.

From the above examination between Norm-Referenced Tests and Standard-Based Assessments, it seems that the nature of Norm-Referencing ultimately dictates that some portion of students must fail for others to succeed. Although these two methods may produce similar results as of now, it is our obligation as educators to ensure that everyone has the opportunity to succeed. By implementing Standards-Based Assessments and with enough effort: "Yes, they can all succeed."

At the end of this paper I would like to call on the scholar community to build upon my work, as the implementation of Standards-Based Assessments can indeed produce highly valuable results for our younger generation. More research still needs to be performed to

determine the technicalities of establishing an effective feedback system, but this is certainly able to be accomplished. Several limitations include the difficulty of aligning all states to accept the same standard, and more funding required because there is a greater demand for more faculty due to the increased workload involved for those who provide extensive feedback.

REFERENCES

- [Aviles 604] Aviles, Christopher B. "Grading With Norm-Referenced Or Criterion-Referenced Measurements: To Curve Or Not To Curve, That Is The Question." *Social Work Education*, 20.5 (2001): 603-608. SocINDEX with Full Text. Web. 27 Feb. 2016.
- [C. Edwards and L. Edwards] Edwards, Clifford H., and Laurie Edwards "Let's End The Grading Game." *Clearing House*, 72.5 (2001): 260. Academic Search Complete. Web. 27 Feb. 2016.
- [Hamp-Lyons] Hamp-Lyons, Liz. "Farewell To Holistic Scoring?" *Clearing House*, 72.5 (2001): 27.(2016): A1-A2. Academic Search Complete. Web. 26 Feb. 2016.
- [Harsch and Martin] Harsch, Claudia, and Guido Martin. "Comparing Holistic And Analytic Scoring Methods: Issues Of Validity And Reliability." *Assessment In Education: Principles, Policy & Practice*, 20.3 (2013): 281-307. Academic Search Complete. Web. 22 Feb. 2016.
- [Hattie and Timperley] Hattie, John and Helen Timperley "The Power of Feedback" *Review of Educational Research*, (2007): 81-112. Academic Search Complete. Web. 22 Feb. 2016.
- [Polloway et al.] Polloway, Edward A., and Michael H. Epstein. "Classroom Grading: A National Survey Of Policies." *Remedial & Special Education*, 15.3 (1994): 162.

- Academic Search Complete. Web. 27 Mar. 2016.
- [Schullery] Schullery, Nancy M. "A Holistic Approach To Grading." *Business Communication Quarterly*, 66.1 (2003): 86-90. Business Source Complete. Web. 21 Feb. 2016.
- [Scriffiny] Scriffiny, Patricia L. "Seven Reasons For Standards-Based Grading." *Educational Leadership*, 66.2 (2008): 70-74. ERIC. Web. 27 Feb. 2016.
- [Sharma] Sharma, Priyanka. "Standards-Based Assessments In The Classroom: A Feasible Approach To Improving The Quality Of Students' Learning." *Contemporary Education Dialogue*, 12.1 (2015): 6-30. ERIC. Web. 26 Feb. 2016.
- [Shippy et al. 16] Shippy, Nicole, Barton A. Washer, and Billie Perrin. "Teaching With The End In Mind: The Role Of Standards-Based Grading." *Journal Of Family And Consumer Sciences*, 105.2 (2013): 14-16. ERIC. Web. 27 Feb. 2016.
- [Swarbrick 4] Swarbrick Nancy. "Primary and secondary education - Standards and examinations" *Te Ara - the Encyclopedia of New Zealand*, (2012). Web. 17 Apr. 2016
- [Tata] Tata, Jasmine. "The Influence Of National Culture On The Perceived Fairness Of Grading Procedures: A Comparison Of The United States And China." *Journal Of Psychology*, 139.5 (2005): 401-412. Business Source Complete. Web. 27 Feb. 2016.
- [Tierney et al.] Tierney, Robin D., Marielle Simon, and Julie Charland. "Being Fair: Teachers' Interpretations Of Principles For Standards-Based Grading." *Educational Forum*, 75.3 (2011): 210-227. ERIC. Web. 21 Feb. 2016.
- [No Child Left Behind] United States. "No Child Left Behind Act of 2001." *Cong. United States Department of Education.*, 107th Cong. Cong. Bill. N.p., n.d. Web. 27 Mar. 2016.
- [U.S. NCES] United States. "Overview and Inventory of State Education Reforms: 1990 to 2000." *National Center for Education Statistics.*, N.p., n.d. Web. 27 Mar. 2016.
- [U.S.] United States. "Aligning Mathematics Assessment Standards: New Mexico and the 2009 National Assessment of Educational Progress (NAEP)." *National Center for Education Statistics.*, N.p., n.d. Web. 27 Mar. 2016.
- [U.S.] United States. "Profile of State Assessment Standards, Alaska." *National Center for Education Statistics.*, N.p., n.d. Web. 27 Mar. 2016.
- [U.S.] United States. "Profile of State Assessment Standards, Delaware" *National Center for Education Statistics.*, N.p., n.d. Web. 27 Mar. 2016.
- [U.S.] United States. "Profile of State Assessment Standards, Michigan." *National Center for Education Statistics.*, N.p., n.d. Web. 27 Mar. 2016.
- [U.S.] United States. "Profile of State Assessment Standards, South Dakota." *National Center for Education Statistics.*, N.p., n.d. Web. 27 Mar. 2016.
- [U.S.] United States. "Table 169. States That Use Criterion- referenced Tests (CRTs) Aligned to State Standards, by Subject Area and Level: 2006-07." *National Center for Education Statistics.*, N.p., n.d. Web. 27 Mar. 2016.
- [U.S.] United States. "Table 2.4. Science Statewide Assessments in Grades 3-8, by State: 2015" *National Center for Education Statistics.*, N.p., n.d. Web. 27 Mar. 2016.